下载APP

16 | 产品体验的评估也可以量化吗？

2021-02-19 炒炒

体验设计案例课 进入课程 >

讲述：炒炒
时长 14:00 大小 12.83M

你好，我是炒炒

在我之前与其他部门的一次协作中，我公司的 CTO 问了我很多问题：

> 我是个典型理工男，不太懂你们这个体验专业，但我习惯用数据来判断。我经常听你们说产品体验好，为什么好呢？可以打几分？这种体验好可以用数据衡量吗？最好的是什么样的？可以打 100 分吗？我投入多少个体验人员是最划算的？如何评估这个投入产出多少价值？

之所以我印象这么深刻，是因为我第一次被 CTO 提了这么多的问题。其实，他问的这些问题总结来说，就是在问一件事，**是否可以用量化的数据来评估产品的体验好坏**。

世界的万事万物都是可以被度量和评估的，我们的用户体验领域自然也不例外。

我们在学习“08 | 怎样搞定一场可用性测试？”的时候曾讲过，我们可以通过用户的可用性测试，使用相对量化的数据来帮助我们发现产品的问题，对产品进行迭代和优化。

这种可用性测试的方法是从**产品视角**，关注用户使用产品能否成功完成某些任务。

而用户体验度量是从**用户视角**，衡量用户使用产品或系统时的个人体验。它是一个更宏观的概念，强调的是用户与产品之间的整体交互，以及交互中形成的想法、感受和感知。

那么这种抽象的东西，真的可以度量吗？如何度量呢？

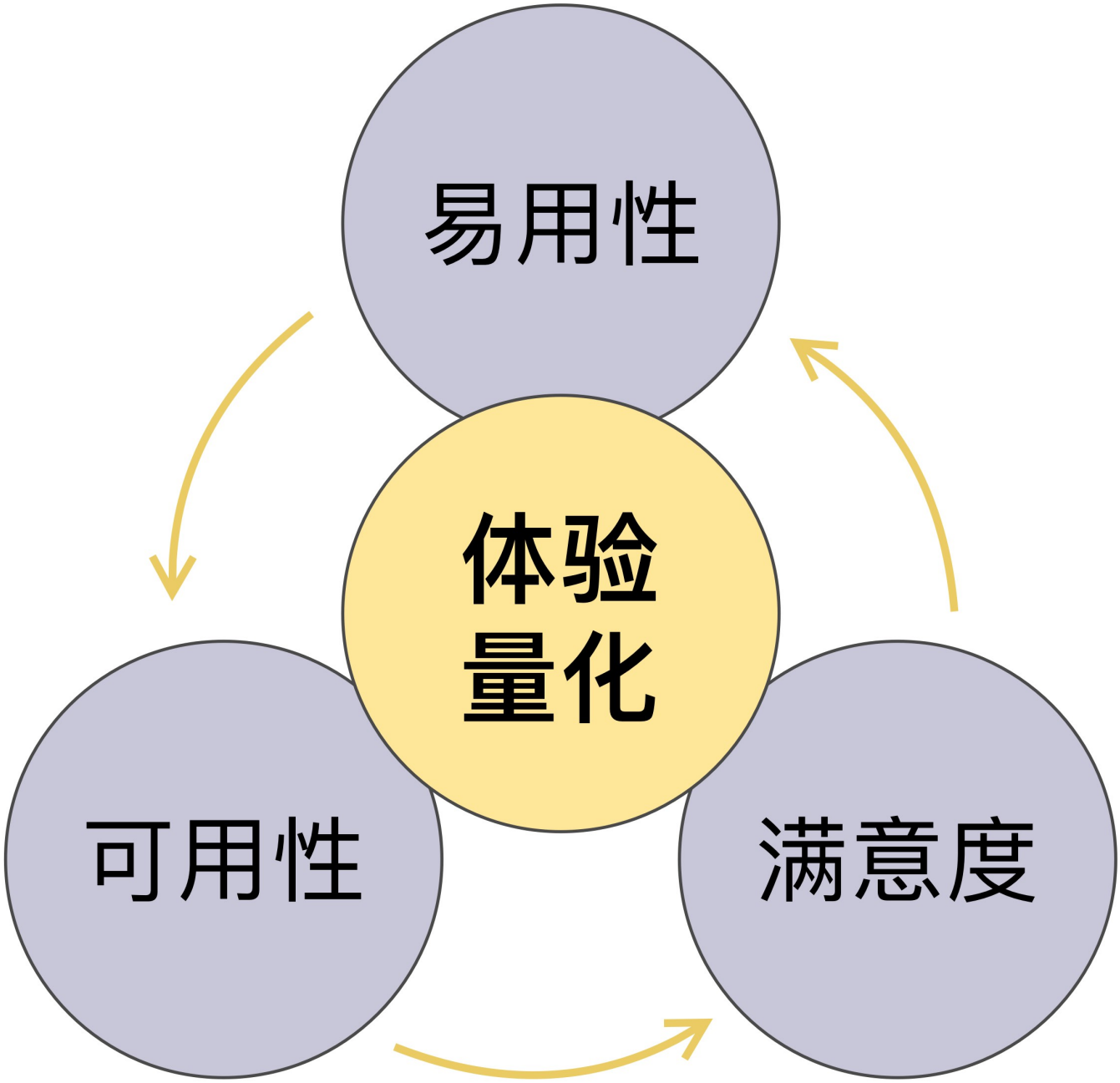
体验能度量吗？

我先给你举两个小例子。

当你用盒马下单后，系统会让你给骑手进行评分。评分体系包含速度、态度、时效等这样的指标，这个其实就是对体验的一个打分收集。打分的人多了，就会生成一个服务评分。平台通过服务评分，就可以知道这个骑手服务哪里好、哪里不好，如何进行结构性优化或者激励。

这个就是体验的量化拆解在日常行为中的一个应用。所以，体验是可以评分的、可以度量的。

大厂一般会有一个组织，专门负责给公司的所有产品进行产品体验的打分，用以帮助该产品进行体验优化。我们一般会从**可用性、易用性、满意度**三个方面去打分，我称之为打分模型。



一般来说，可用性看的是产品的基础功能是否过关，如果不过关的话，很有可能会影响到用户对产品的第一体验；而易用性更多地代表一个产品好，好到什么程度，差的话，又差成什么样子。

而满意度偏向于用户个人主观的一个感受，代表的是一种整体上给用户带来的感觉。这三个维度共同决定了一个产品给用户带来的体验的好坏。

那么，具体该如何利用打分模型去度量体验呢？主要分为五个步骤。

如何度量体验？

第一步：定范围

之前我有说过，我最近两年一直在做 B 端产品，所以 C 端的那种大范围随机发调研问卷或者走廊测试这样的搞法，对 B 端来说，性价比不是很高。B 端系统大、复杂度高、模块多、岗位角色多 ，所以一般会先定范围。

就拿我们之前的企业网银这款产品来说，这个企业网银集合了五大业务中心的业务，相关功能模块有 400 个左右，涉及到的页面有 3000 多个。如果我们要对整个企业网银进行整体打分，那这个工作量太大了，而且还要协同不同的业务中心，整体推进起来的速度也会非常缓慢。

所以，我们根据不同的业务属性划分成了不同的范围，比如说转账、资金池、票据管理、跨境中心、理财等。

第二步：选竞品

这里你可以回忆一下我们的“04 | 为什么你的竞品分析看起来像拼接抄袭？”中讲到的，我们选竞品的方法。基于现有的目的，我们一般选**直接竞品**就可以了。比如说，我们要评估的是财资系统，我们会直接挑选行业里做的最好的财资系统，来进行模块的详细对比。

结合我们在上面提到的打分模型，我们可以把可用性、易用性和满意度对应我们的产品，做一个拆分。比如说，可用性分为是否满足用户需求、流畅性怎样、文案描述是否准确等等。

第三步：定规则

接下来，对于每个细项指标的评分，你可以采用五级量表的 10 分制评分。你可以根据自己的操作体验，给每一个细项进行评分。比如说，评分对应规则如下：非常不同意（1、2 分）、不同意（3、4 分）、一般同意（5、6 分）、比较同意（7、8 分）、非常同意（9、10 分）。

比如说，如果你要转账，结果使出了吃奶的力气，转账都完成不了，那这个明显是“非常不同意”，1 分都觉得多了，对吗？所以，就拿“满足用户需求”这个指标来衡量的话：

> 如果你觉得该产品提供的相关功能基本都能够满足你的需求，那你就可以选择“同意”或者“非常同意”对应的分值；

如果这个产品只是满足了你的部分需求，你对该产品的态度既不喜欢，也不排斥，那你可以选“一般”这个分值；

如果这个产品比较糟糕，不能够满足你的核心需求，那你就可以选择“非常不同意”或“不同意”对应的分值；

举个例子，假设你现在需要在 app 上进行转账，结果使出了吃奶的力气，转账都完成不了，那这个操作体验在“满足用户需求”这个指标上，明显是“非常不同意”，1 分都觉得多了，对吗？

我总结了一个表格，给你列举了一下我在最近的一个项目中用到的产品体验分评估指标。这个表格针对的是一个金融类的 B 端产品，应该会给你一些更具象的参考。

当然，在工作中，你需要根据你所服务产品的实际情况不同进行指标的增减。

评分维度	评分指标	指标说明	分值1–10
可	满足用户需求	能够满足多种业务场景，帮助用户完成多项工作。在当前系统不能满足用户需求时，可以给用户一些明确指引，帮助用户利用其他系统完成工作	
	角色/功能/权限划分清晰	功能结构划分清晰，用户能够清晰的掌握相关模块功能；不同角色与功能一一对应，避免出现角色和功能对应混乱的情况，增加用户理解成本	
	安全及隐私	响应用户的安全和隐私诉求，在输入密码的场景使用安全键盘，在资金相关的页面给予用户相应的安全提示，避免收集不必要的用户个人	

可用性		提示，避免收集不必要的用户个人信息	
	系统的流畅性	操作过程中无卡顿，新页面的加载时长不超过2s，不会出现崩溃或系统卡死的情况。因系统问题操作流程中断的，再次启动是否恢复到断点页面	
	导航结构清晰	结构分类清晰，方便用户快速查找到相关功能；用户操作路径清晰，用户可以清晰的找到自己要去的页面，并且能够轻易的沿着原路径返回	
	文案描述的准确性	对于行业术语或指标能够用通俗易懂的方式描述清楚，并且前后保持一致；提示信息或者规则应该准确，避免存在歧义	
	……	……	
	状态可见性	保证界面上内容可见，让用户对当前页面的信息有全面的掌握；对于页面控件的可用状态、不可用状态有明显的区分；及时对用户的操作给予反馈，有了操作结果（成功/失败）及时告知用户，等待中尽量告知用户等待时间	
	一致性	所有的操作均符合用户预期，类似功能的操作保持一致，符合用户操	

易用性	易学性	作习惯，对于一些复杂的操作给用户适当的引导，降低用户学习成本	
	容错性	在用户操作遇到错误的时候对用户行为进行引导，使用户可自由的进入或退出	
	信息的易读性	信息分布简洁清晰，重点内容突出，方便用户快速浏览，避免因为功能和信息的繁杂影响用户的操作。对于需要用户重点关注或者必须关注的信息有对应的提示。不使用过度的设计而影响到用户的信息阅读	
	帮助和提示	对于需要提示的信息提前告知用户，避免用户操作后产生了疑问再提示用户；在用户遇到问题的情况下能够提供及时的帮助和引导	
	场景贴合性	信息展示及相关功能符合用户特征及所在场景，避免引起用户的困惑和反感	
	控件/字体/功能色彩的一致性	不同页面间的控件使用、字体层级、功能色的使用要保持统一，避免因为不统一造成用户的不适或错误操作	
	icon的识别	icon的设计应该符合整体视觉风格，并且图形的表意准确，符合用	

	的识别性	恰，并且图形的表意准确，符合用户对图形的心理预期	
	……	……	
满意度	品牌元素的一致性	对于图片风格和色彩风格的使用要符合产品整体的品牌风格，避免因为风格的不一致影响整体风格的协调性，对品牌建设产生消极影响	
	……	……	

PS：你可以根据自己的产品情况，设计自己的体验打分表。

你可以根据自己的产品情况，设计自己的体验打分表。

第四步：走查打分

接下来，我们一般的做法，就是根据标准分模块，按照业务流程去走查打分。

为了保证最终打分的合理性，我们会在打分前给各位打分人**做一次宣讲**，把打分规则详细地讲一遍，如果有疑问的地方让大家现场讨论，这样可以使各位打分人对规则的理解更趋于一致。

对齐打分规则后，打分人对本次需要打分的模块进行依次打分，且彼此之间互不影响。比如，本次划定的打分范围是模块 A、模块 B、模块 C 三个模块，则每个打分人需要依次完成对这三个模块的打分。

这里要求走查的人员**对业务要很熟悉**，也就是需要基于懂业务的体验走查。不然发现的问题都是很浅层次的视觉问题。倘若发现的都是这样的问题，对于一个业务系统来说，就是**形式大于逻辑**了，收获不到对体验提升有决定影响力的、特别有价值的信息。

举个例子，我们最近做某财资平台的体验打分，发现所有的打分都是视觉表现层的一些扣分项，例如像素没对齐，弹框样式不统一这样的不疼不痒的问题，这对产品的体验优化起不了决定性的帮助，同时也显得体验问题都集中在视觉层面，应该由视觉设计师对体验问题负责。

溯源原因，就是因为走查打分的人员里没有加入业务人员，大家不懂业务流程，所以只能按照产品说明书点击，很难走进产品真实的使用场景里面去，很难发现产品真实的体验问题。

第五步：输出体验分值与建议

最后一步，就是输出体验分值和建议

通过以上流程的打分，我们就能看到每个评分指标的得分，那么怎么计算一个最终的体验分呢？

我们的计算逻辑是这样的：先把每个评分指标的分数相加，再除以评分指标个数，就能得到单个用户对该模块的体验分；然后把每个用户对该模块的评分相加，再除以参与评分的用户数，就得到了该模块的体验分。

$$\text{体验分} = \frac{\text{用户A 评分=}\ \text{细项指标分数和/指标个数} + \text{用户B 评分=}\ \text{细项指标分数和/指标个数} + \cdots}{\text{用户数}}$$

计算出来了体验分，你可能会有一个疑惑，在我们列举的打分指标中，为什么易用性的权重这么大，占了 8 个指标，满意度却只有 1 个指标？

这也是我正要提醒你的，我列举的这个表是针对我们自己的 B 端产品使用的评分表的局部，我们在体验评估中更偏重产品的易用性，所以会把易用性的指标权重设置的大一点。

在实际工作中，你可以结合产品的实际情况去调节可用性、易用性、满意度三个维度的权重，这样得出的体验分才更符合产品的需要。

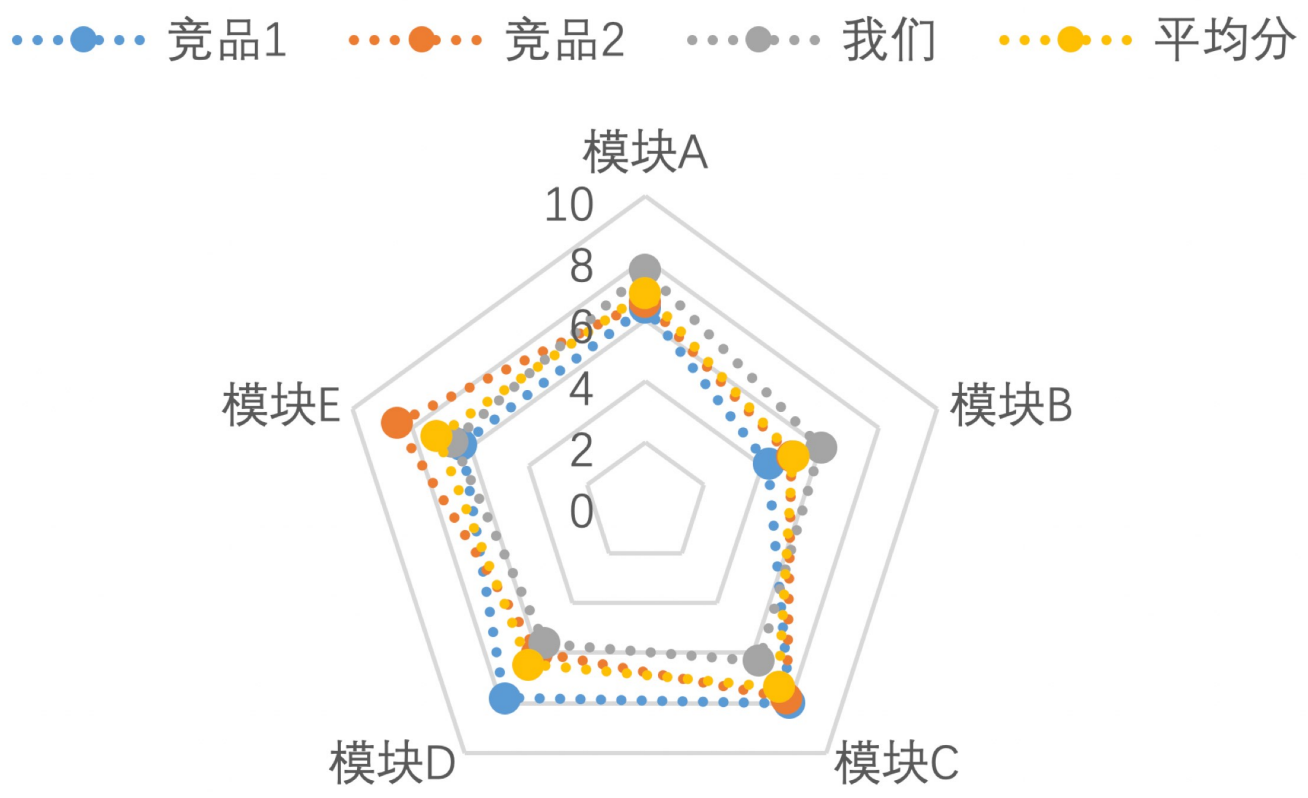
当然，除了看总体的体验分，我们还可以对比看单一维度的体验分，比如可用性或易用性的体验分。数就是这么些数据，我们可以根据我们的需要灵活运用。

当每个模块的体验分计算出来后，我们又该怎样去对比分析呢？

通常情况下，我们会参考两个标准进行对比：**行业标准**和**自我标准。**

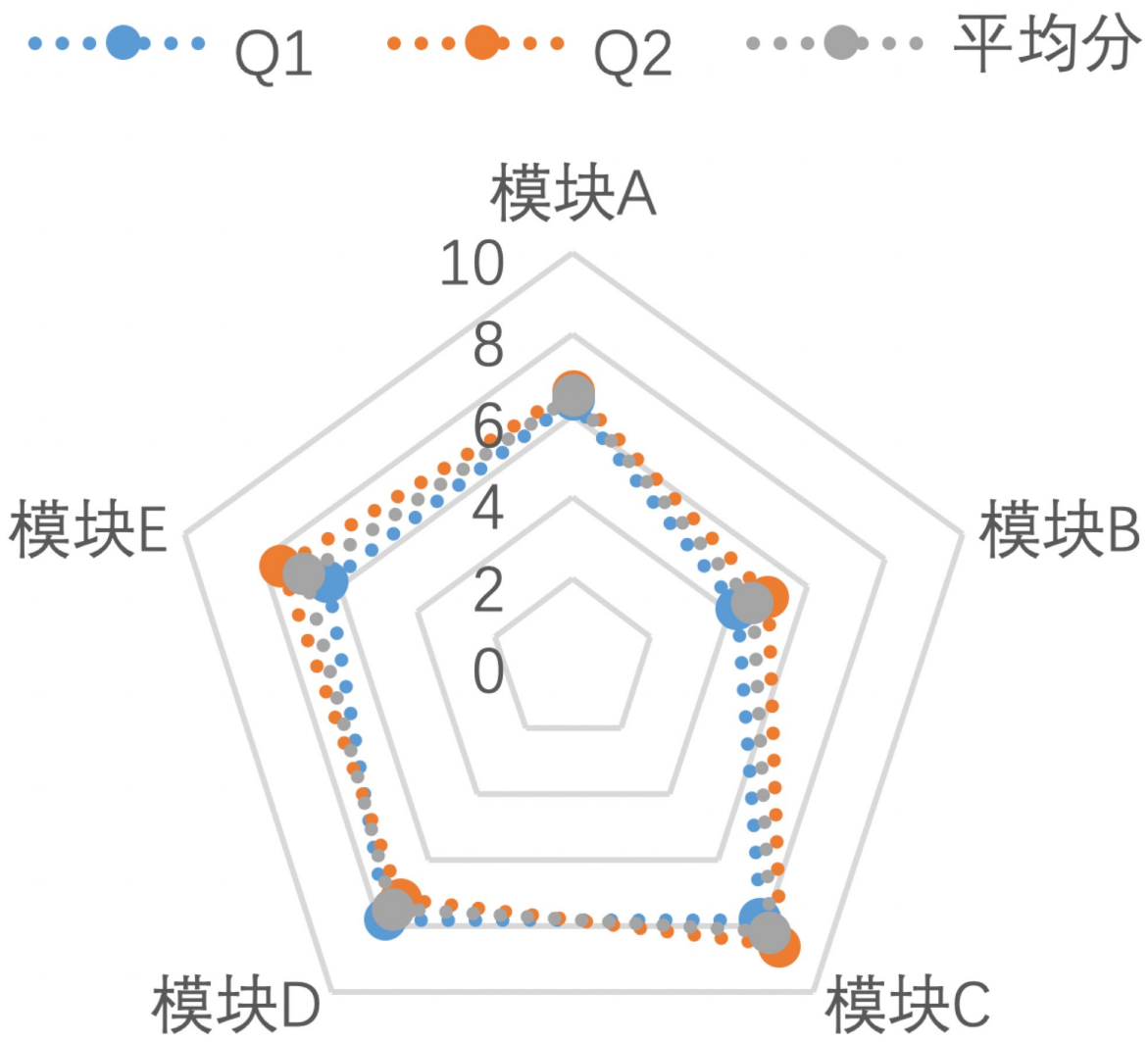
先来看行业标准，对于很多 B 端的产品来讲，可能你在市场上找不到比较通用的行业标准。那怎么办呢？你还记得在第二步中讲到的选竞品吗？假设我们选了两个竞品，那么参与评分的用户，同时会对竞品进行评分，这样，我们也用同样的逻辑计算出竞品的体验分。

再计算两个竞品的平均分，我们就可以简单粗暴地把两个竞品的平均分定义为行业标准。这样，我们就可以跟行业标准对比。倘若低于行业标准，那说明我们做的太烂了，需要深度优化产品的体验。



再来看一下自我标准。互联网产品追求的是每一次都有进步，而不是完美。这就要求我们的团队要不断地对产品进行迭代优化。所以，这个体验度量的工作不是一次性的，也是迭代的。

那么，我们可以跟上个季度比，看看我们的分值有升高吗？没有升高的话也需要反思，是哪个模块的哪个要素分值特别低，我们就可以有针对性地进行优化。



通过体验分值来辅助产品进行优化决策，这样的决策才更有方向性和目标性。

体验度量的价值

通过度量，我们可以量化地知道，产品的体验好不好，好在哪儿？不好在哪儿？跟行业相比，我们是处于领先地位还是追赶者地位？用户体验度量的最终目标不在于度量本身，度量只是一种途径或方法，可以帮助你获取更多信息，以便做出更好的决策。

用户体验度量可以回答很多对组织来说至关重要的问题，以及其他方法回答不了的问题。比如：

用户会非常认同我们产品吗？

这个新产品的使用效率会高于当前的产品吗？

与竞品相比，这个产品的用户体验如何？

这个产品中最明显的可用性问题是什么？

……

有了体验度量，我们可以确定很多抽象的问题，比如说，一个简单的改变可以带来多少具体的价值、从哪个方向降低完成客户服务任务所需要的时间、增加每天处理的多少交易量，可以使未完成的客户订单量减少多少、如何提升客户满意度和增加交易等等。

当我们的评估**从总体上给企业带来收入上的增加**时，体验度量就有了价值。

炒炒总结

今天我们讲到了产品体验的评估问题，讨论了是否可以利用量化的手段给体验评分。

首先，我们确定了，产品体验是可以评估和度量的。我们在对产品进行用户体验度量的时候，可以从可用性、易用性、满意度三个大的维度去评估，建立我们的打分模型。

第一步，先划定好范围。我们的打分对象可以是针对一个完整的产品，也可以是产品中的某个相对独立的功能模块；第二步，按照我们之前所学，选择好我们的直接竞品；

第三步，结合产品属性和评分目标，进一步细化打分指标，每个细项指标的打分规则可以采用五级量表 10 分制的模式；第四步，打分的时候，要保证评分人的独立评分，并且要确保打分的人熟悉业务场景，尤其是针对 B 端的产品，在熟悉业务场景的基础上，进行打分会更有客观业务价值；

最后，在对体验分的结果进行分析对比的时候，我们可以根据需要，从行业标准和自我标准两个层面来对比分析。体验度量的价值并不是打分，而是通过评出的体验分让产品的关联方对产品的体验有更客观的认知，同时帮助产品更有效地完成迭代优化，带来业务上的增长。

希望通过本节课的讲解，帮助你在工作中更客观地评估产品的用户体验。

课后题

请结合课程中讲的详细评分维度，给你现在服务的产品或者重点功能进行一个评分吧。

记得在留言区和我讨论、交流你的想法，每一次思考都会成为你进步的基石。

如果你喜欢今天的内容，也欢迎你把这一讲分享给你的朋友。

感谢你的阅读，我们下一讲再见。

提建议

© 版权归极客邦科技所有，未经许可不得传播售卖。 页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇 15 | 如何建立设计方案的验证模型?

下一篇 17 | 如何轻松应对不同产品阶段的设计量化?

精选留言 (3)

写留言

Wayne
2021-03-12

评分指标是怎么定出来的呢

展开

王晓冰
2021-03-02

如果从找体验问题、求产品进步的角度来看，炒炒老师的实践真的很棒，也很接地气。想请教下炒炒老师，业内有很多度量模型，包括Google的heart模型、蚂蚁的ptech模型，这些模型大多用到了行为数据来度量一个平台的体验水平，感觉是相对系统化的运作模型。想听下您对量化体验的有效性（可能包括假设检验）、加入行为数据测量等方面的看法?

展开

作者回复: 利用用户行为数据作为元素来评估产品的体验水平是比较通用的方法，毕竟用户体验的好坏会直接影响用户的行为。当然了，对于一些商业化比较明显的产品，也会适当的加入一些商业数据指标的考量，这些都是结合自家产品的不同情况来确认的。

你可以接着往下来17课，里面更详细的讲了在不同产品阶段的一些评估方式，或许能给你带来更多的启发。

石云升
2021-02-19

我理解这个量化的本质是，当所有人的主观感受通过指标汇集之后，就会得出客观的结果。
所有人的主观汇总就是客观。

展开

作者回复: 哈哈，你这个逻辑也说得通。